Tutorial

Advances in Methods and Practices in Psychological Science
April-June 2026, Vol. 9, No. 2,
pp. 1–20

DOI: 10.1177/25152459261419036
www.psychologicalscience.org/AMPPS

The Language-Based Assessment Model Library: Open Model Sharing for Independent Validation and Broader Applications

August H. Nilsson[1], Veerle C. Eijsbroek[2], Zhuojun Gu[2], Katarina Kjell[2], Salvatore Giorgi[3], Roman Kotov[4], Adithya V. Ganesan[4], H. Andrew Schwartz[5,6], and Oscar N. E. Kjell[2,5,6]

[1]Oslo Business School, Oslo Metropolitan University, Oslo, Norway; [2]Department of Psychology, Lund University, Lund, Sweden; [3]Department of Computer and Information Science, University of Pennsylvania, Philadelphia, Pennsylvania; [4]Department of Psychiatry and Behavioral Health, Stony Brook University, Stony Brook, New York; and [5]Department of Computer Science, Stony Brook University, Stony Brook, New York; [6]Department of Computer Science, Vanderbilt University, Nashville, Tennessee

Abstract

Language-based assessments (LBAs), quantitative estimates of scientific constructs based on language, have advanced methods in the psychological and social sciences for more than a decade. LBAs based on individuals' prompted descriptions analyzed with large language models to produce scores of their psychological states and traits have shown strong convergence with the corresponding rating scales ($r > .80$) and have often surpassed rating scales in predicting theoretically relevant behaviors (external criteria). Despite their high validity across numerous psychological outcomes and contexts, the broader adoption of LBA models (LBAMs) has been limited. Even when made available alongside research publications, these models often remain inaccessible because of technical complexities, inconsistent documentation, and the absence of a standardized repository. In this tutorial, we introduce a framework targeted to social and psychological scientists for accessible sharing models with others—the Language-Based Assessment Models (L-BAM) Library—and a toolkit for easily using LBAMs via the *text* package in R. L-BAM covers a wide range of models for assessing mental-health disorders (e.g., depression, anxiety), well-being (e.g., satisfaction with life, harmony in life), implicit motives (need for power, affiliation, and achievement), and more. The L-BAM Library aims to increase the availability and resource efficiency of LBAs of psychological constructs while encouraging replication, independent validation, and the broad application of preexisting LBAMs.

Keywords

language, artificial intelligence, open data, open materials, language-based assessment models

Received 5/1/25; revision accepted 12/24/25

The language people use to describe themselves and their state of mind can answer research questions concerning what they think (Al-Mosaiwi & Johnstone, 2018), how they feel (Pennebaker et al., 2003; Zimmermann et al., 2017), what they do (Hu et al., 2016), who they are (e.g., J. Chen et al., 2020; Kwantes et al., 2016), how they interact with others (Bayram & Ta, 2019; Ireland

Corresponding Authors:
August H. Nilsson, Oslo Business School, Oslo Metropolitan University, Oslo, Norway
Email: august.nilsson1907@gmail.com

Oscar N. E. Kjell, Department of Psychology, Lund University, Lund, Sweden
Email: oscar.kjell@psy.lu.se

et al., 2011), how they make sense of the world (Fausey & Boroditsky, 2010; Sterling et al., 2020), how they behave (Mehl et al., 2007; Tidwell et al., 2025), and much more. Language provides rich psychological information that extends beyond traditional closed-ended assessment methods (Boyd et al., 2024; Kjell, Kjell, & Schwartz, 2024).

Language-based assessments (LBAs) can be viewed as a new family of psychological-measurement tools based on the assumption that language reliably reflects underlying states, traits, values, thoughts, feelings, and so on (Boyd et al., 2024; Kjell et al., 2019; Park et al., 2015). Unlike traditional closed-ended scales, which constrain responses to predefined items, the LBAs leverage the natural expressiveness of language to derive quantitative assessment scores from natural language. For example, a large language model can be used to convert natural language (e.g., social media posts) into numerical representations, which are then used as predictors in a regression model to estimate depression-severity scores. Such models are called "LBA models" (LBAMs; see Argamon et al., 2007; Boyd & Schwartz, 2021; Kjell et al., 2024; Park et al., 2015; Tausczik & Pennebaker, 2010). They may serve as a complementary method to traditional assessment methods, such as informant reports, behavioral tasks, or physiological recordings, but with unique advantages tied to linguistic richness.

Using language to quantitatively assess psychological states and traits offers several advantages. First, natural language is the primary means through which individuals express complex psychological experiences (e.g., Tausczik & Pennebaker, 2010). Second, natural language possesses great measurement properties, including, for example, broad range, fine resolution, and openness (Kjell et al., 2024). The broad range of language (e.g., close to a million words in English vs. five, seven, or 11 scale steps) allows individuals to express extreme states (e.g., from hopeless to ecstatic), and its fine resolution enables distinctions between subtle emotional nuances (e.g., differentiating between worried, uneasy, tense, and panicked). Finally, the openness of language enables individuals to generate personalized responses, overcoming the limitations of predefined response categories found in traditional assessments.

Over the past years, researchers have made use of language quantitatively by transforming language to numbers using, for example, large language models, and with the numbers, developing regression models to predict psychological outcomes. These models make use of the nuances of language to predict a certain criterion, and these models are called "LBAMs" (see Argamon et al., 2007; Boyd & Schwartz, 2021; Kjell et al., 2024; Park et al., 2015; Tausczik & Pennebaker, 2010). Recently, several LBAMs have been developed and validated to assess psychological constructs, such as depression severity (Gu et al., 2024), harmony in life (Kjell et al., 2022), and implicit motives (Nilsson et al., 2025). Despite their validity across numerous psychological outcomes and contexts, the broader adoption of LBAMs has been limited. In this tutorial, we introduce and describe the Language-Based Assessment Model (L-BAM) Library, which serves as an open library for sharing pretrained LBAMs in which the models can easily be used with one function from the R package *text*. The L-BAM Library aims to facilitate the reproducibility, comparability, and accessibility of LBAMs by providing standardized tools and methodologies for researchers and is targeted toward social and psychological scientists. By making these models easily available, we encourage independent validation, broader application across diverse psychological domains, and more efficient use of existing resources. In this tutorial, we outline how researchers can use LBAMs to assess psychological constructs and provide guidelines for contributing new models to the library.

LBAs Can Improve Psychological Science

Accurate quantification of mental states and traits is essential for psychological science, enabling researchers to systematically assess, compare, and track psychological constructs and experiences across individuals. Over the last 90 years, rating scales based on narrowly defined questions coupled with closed-ended questions (i.e., Likert scales; Likert, 1932) have come to dominate the assessment of psychological constructs. Although the rating-scale method has led to important findings, the format comes with limitations, such as constraining respondents to comprehensively describe their unique experiences and state of mind. Although language initially is more complex to analyze than rating scales, recent advancements in artificial intelligence (AI) and natural language processing now allow researchers to translate rich language descriptions into meaningful numerical assessments that align with and even enhance traditional psychometric measures.

Methodological flexibility

LBAs offer considerable methodological flexibility and have been increasingly applied across a wide range of psychological constructs and related behaviors (e.g., Boyd & Schwartz, 2021; Kjell, Kjell, & Schwartz, 2024; Mihalcea et al., 2024). LBAs have enabled researchers to assess, among others, personality (Park et al., 2015; Schwartz et al., 2013), implicit motives (Brede et al., 2025; Nilsson et al., 2025), well-being (Jaidka et al., 2020; Sametoglu et al., 2024), and mental illness, such as depression (Gu et al., 2025; Perlis et al., 2024), anxiety (Gu et al., 2025; Teferra & Rose, 2023), and posttraumatic stress disorder (Son et al., 2020). LBAs can also be developed for behaviors that are theoretically relevant to psychological constructs, such as alcohol consumption (Jose et al., 2022; Nilsson et al., 2024), cooperation (Kjell et al., 2021), and suicide (Y. Chen et al., 2024; W. Zhou et al., 2023); somatic diseases (e.g., heart disease, Eichstaedt et al., 2015; or cancer, S. Zhou et al., 2022); and demographic variables, including age and gender (Ganesan et al., 2021; Sarwar et al., 2024).

LBAs can be applied to both probed language data—elicited through targeted open-ended questions—and already existing language data gathered from natural contexts. Probed LBAs ask individuals to describe their state of mind, personal experiences, or specific topics in their own words. These assessments have demonstrated very strong convergent validity with traditional rating scales, with an accuracy approaching or reaching the scales' reliability, which is the theoretical upper limit of concurrent accuracy (r > .80; Gu et al., 2024; Kjell et al., 2022; Nilsson et al., in review). In addition, there are several examples of using probed language (i.e., answers to targeted open-ended questions), which includes asking participants to describe their activities (Nilsson et al., 2022) and themselves in various ways (e.g., Kwantes et al., 2016), recalling various memories (Yeung et al., 2024), reporting stream of consciousness (Sripada & Taxali, 2020), and so on.

LBAs using already existing language data have demonstrated the ability to assess a wide range of physical and psychological outcomes. For instance, social media language has been linked to mental- and physical-health markers (Eichstaedt et al., 2015, 2018; Kjell, Giorgi, Schwartz, & Eichstaedt, 2023), and transcripts of everyday speech have been used to capture emotional fluctuations throughout the day (Sun et al., 2020). There are many existing sources of language that can be analyzed using LBAs, including chats, blogs, text messages, emails, letters, personal diaries, and song lyrics. In addition, language from more specialized settings, such as therapy-session transcripts (Lalk et al., 2024), medical notes (Shah, 2024), and political speeches (Liu, Zhang, et al., 2022), can offer valuable insights for psychological analysis. Together, these two methodological approaches—probed and naturalistic—allow researchers to tailor LBAs to a wide range of research designs and data sources.

Theoretical depth

LBAs also hold promise for advancing psychological theory. For instance, LBAs can be used to explore how individuals naturally express psychological phenomena, offering bottom-up insights that can refine existing theories about constructs (see e.g., Bucur et al., 2021; Coppersmith et al., 2014; Gu et al., 2025; Liu, Ungar, et al., 2022; Nilsson et al., 2024; Stade et al., 2023). Furthermore, because LBAMs can be applied on a large scale quite easily, it opens the door for new applications that can help expand research in a field. For example, implicit-motive (i.e., subconscious needs) assessments have historically been resource-intensive because coding text for motives requires a lot of time and brain power, limiting research and theory development. With automated coding from implicit-motive LBAMs (e.g., Brede et al., 2025; Nilsson et al., 2025), it is possible to assess implicit motives at a much larger scale than was ever practically possible before: in terms of both assessing implicit motives via the classic picture-story exercise and applying the implicit-motive LBAMs on other texts, such as company reports or social media texts (which the coding manual theoretically allows; Winter, 1991), to understand, for example, if power-oriented companies are more or less successful.

By aligning psychological measurement with natural human expression, LBAs enable individuals to communicate their experiences in their own words, offering a powerful complement to traditional closed-ended assessments. For an overview of research studies developing LBAs included in the L-BAM Library, see Table 1.

The Need for the L-BAM Library

Despite solid evidence for the broad applicability, usefulness, validity, and reliability of LBAs, the sharing of models so that they can be easily used by others is currently limited, and there is no centralized library facilitating information and model sharing. Even when made available alongside research publications, these models often remain inaccessible because of technical complexities, inconsistent documentation, and the absence of a standardized library. These limitations restrict resource efficiency, hinder replicability, and impede independent evaluation and systematic testing of generalizability. We believe sharing LBAMs is essential for five key reasons: (a) It is resource-efficient to share because not every research group needs to develop their own models; (b) it supports replication and independent validation, which are critical for tackling psychology's replication crisis (Simmons et al., 2011); (c) it ensures increased comparability when the same models are applied across studies; and (d) concerns about generalizability can be systematically addressed when models are openly shared, enabling researchers to evaluate performance across diverse samples, languages, and settings using the same tools. Finally, (e) for LBAs to have a practical impact on psychology (or other fields, e.g., medicine),

Table 1. Example Uses of Language-Based Assessment Models

Constructs	Description	Reference
Mental health		
Anxiety Depression	Individuals described their anxiety and depression using different formats (e.g., selecting words from lists or writing descriptive words, phrases, or texts). These responses were used to train models that predict corresponding rating scales. The models, trained on a development data set ($N = 963$) and preregistered before testing on a prospective sample ($N = 145$), demonstrated high convergent validity (rs = .60–.79).	Gu et al. (2024)
Suicidal risk Self-harm risk	Individuals described any suicidal ideation or self-harm using open-ended text responses. Language-based models, preregistered and tested against best-estimate-assessed outcomes, showed moderate to strong correlations. For suicidality risk, the model achieved r = .57 (disattenuated r = .73). The self-harm model produced a Pearson correlation of .65, accompanied by a disattenuated correlation of .89.	Gu et al. (in progress)
Mental-health recommendation	A language-based assessment model was developed to provide mental-health recommendations on a scale from 1 to 5 in which higher scores indicate greater need for psychological support. The model was evaluated against the best-estimate assessed recommendation made by experts ($N = 212$). The model demonstrated strong criterion validity (r = .82) and convergent validity with clinical scales ($\|r$s$\|$ = .62–.77). Language inputs included descriptions of general mental health, suicidal thoughts, medical history, and selected depression-related words.	Wiebel et al. (in progress)
General mental health	Transcribed language from automated clinical interviews was used to assess general mental health. The preregistered language-based model, trained on a development data set ($N = 1{,}270$), achieved a correlation of r = .35 in the prospective sample ($N = 272$), exceeding the preregistered cutoff ($r > .315$). Performance was substantially better than demographic-only models (r = .13), and adding demographics did not improve predictive accuracy.	Kjell et al. (in progress)
Physical health		
General physical health	The preregistered language-based model assessing physical health (from study above) achieved a correlation of r = .38 in the prospective sample, surpassing the preregistered threshold ($r > .348$) and aligning with development performance. It outperformed models using only demographics (r = .16), and there was no significant gain from including demographic variables alongside language.	Kjell et al. (in progress)
Well-being		
Psychological and subjective well-being	Participants responded verbally or in writing about life satisfaction and autonomy. Language-based models using contextual word embeddings significantly converged with corresponding questionnaire measures (rs = .16–.63). Although satisfaction with life was reliably assessed, autonomy was less predictable.	Mesquiti et al. (2025)
Cognitive well-being	Respondents were prompted to describe their harmony in life and satisfaction with life with various response formats. The best-performing models converged with a corresponding rating scale at r = .85 and r = .80, respectively.	Kjell et al. (2022)
Experienced well-being	Individuals described current feelings across multiple days ($M = 20$) with open-ended responses. Concatenated language predicted average valence and arousal ratings at r = .82 and r = .43, respectively.	Nilsson et al. (in review)
Work well-being	Respondents were prompted to describe their work engagement and job satisfaction. The responses converged with a corresponding rating scale at r = .71 and r = .68, respectively.	Nilsson et al. (in review)
Personality		
Implicit motives		
Power Achievement Affiliation	Using 85,000 sentences from picture-story exercises that were coded for the need for power, achievement, and affiliation, the best performing models converged with the human codings at intraclass correlation coefficients of .90, .88, and .92, respectively.	Nilsson et al. (2025)

models need to be shared in an accessible format that allows the broader scientific community to implement them effectively. Just as researchers have successfully shared validated questionnaires, they can also share LBAMs. Although there are repositories for uploading models, such as GitHub, Hugging Face, and OSF, the L-BAM Library is just that, a library from which the actual models are hosted on repository platforms.

Tutorial

In this tutorial, we use the *text* package (Version 1.8), an R package that lets users download and use large language models and develop LBAMs (Kjell, Giorgi, & Schwartz, 2023). There exist other packages for advanced language analysis, such as *DLATK* (Schwartz et al., 2017), *Keras* (Chollet et al., 2015), and *PyTorch* (Paszke et al., 2019) in Python, but in this tutorial, we focus on the *text* package in R, which streamlines these analyses in a user-friendly way tailored for social and behavioral scientists. In this tutorial, we aim to assist in increasing the sharing of LBAMs by introducing two resources. First, we describe how the *textAssess* function can automatically download models, preprocess language data, and apply models for assessment, prediction, or classification. Second, we introduce the L-BAM Library, where researchers can discover existing models and describe their own models with instructions for how they can be downloaded, used, and cited. We encourage researchers to contribute new models to the library, promoting collaboration and advancing open science in language-based analysis. In the tutorial, we predominantly cover prediction-based models through language. For researchers interested in interpretability and theory-driven exploration based on language analysis, we have developed a complementary tutorial describing multiple methods for visualizing human language (Eijsbroek et al., 2026, under review), which introduces methods such as keyword extraction, topic modeling, and AI-based visualizations that can be used alongside LBAs to provide further psychological insights. Before diving into the tutorial, we want to emphasize some caveats about generalization.

Essential caveat about generalization

LBAs can be developed in one setting (e.g., social media) and applied in another (e.g., clinical interview). Often, however, they do not generalize across all contexts: A model's generalizability depends on several factors, such as the setting, population, distribution of target psychological measure, and language domain (i.e., the similarity between the training-data language and the target language being assessed). Therefore, users must take responsibility for evaluating the appropriateness of each model for their specific context. This includes assessing whether the model's training and evaluation contexts and language distributions are sufficiently similar to their target data (for more information, see the Supplemental Material available online) and if needed, validating the model's performance on a subset of their own data before drawing any substantive conclusions. This is why it is essential to carefully describe each model—its training data, performance metrics, and so on—as outlined in the section The L-BAM Library: Reproducibility, Replication, and Generalizability. Comprehensive documentation helps users evaluate whether a model is appropriate for their specific context and supports transparent, reproducible science. We discuss this further in the Responsible Applications and Generalizability section.

Theoretical overview of the L-BAM Library phases

There are three core phases involved in training LBAMs and applying models from the L-BAM Library. First, the language is converted into numerical representations (i.e., word embeddings) using a large language model (Fig. 1a); second, these word embeddings are used to train a model to assess or predict a criterion variable (Fig. 1b); and third, these models can be applied on new data for assessment or classification (Fig. 1c). Details on how to transform text to word embeddings and train LBAMs using the *text* package have been described in detail before (see Kjell, Giorgi, & Schwartz, 2023). Below, we provide an overview of these two steps (Figs. 1a and 1b) before describing the application and use of the L-BAM Library (Fig. 1c).

Word embeddings. Language can be represented numerically through word embeddings, which are lists of values that capture the latent meaning of words in a structured format. Essentially, the word-embedding process transforms language into numerical representations, making it possible to analyze linguistic patterns computationally. This transformation is powered by large language models, such as GPT-4 or BERT, which are trained on vast amounts of text data from the internet, books, and other sources to develop a generalizable representation of language. The large language models' main task in training is to predict the next word based on the previous context. Through this vast training procedure, the models create a multidimensional semantic space in which words—and even entire texts—can be positioned based on their contextualized meaning and usage such that each model represents language in several slightly different versions (i.e., layers; for more details, see Devlin et al., 2019; Vaswani et al., 2017). Text can be transformed into word embeddings

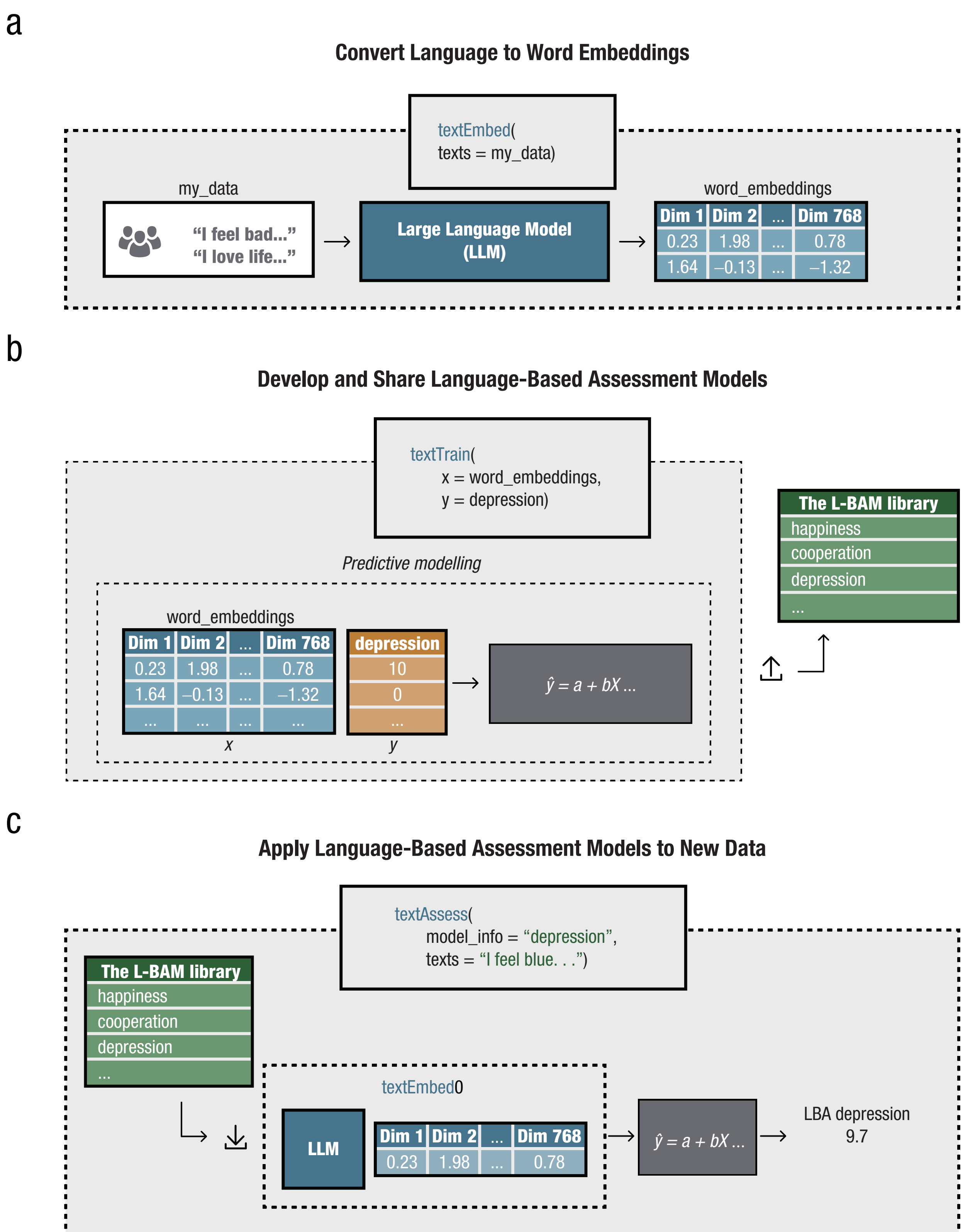


Fig. 1. Overview of the Language-Based Assessment Model (L-BAM) Library components. (a) Convert language to word embeddings. (b) Develop and share language-based assessment models. (c) Apply language-based assessment models to new data.

using the *textEmbed* function from the text package, as described in detail by Kjell et al. (2023). Thus, this function transforms raw language data into meaningful numbers representing the language.

Training LBAMs. With word embeddings, it is possible to create linear regression models to predict psychological constructs and relevant outcomes. Normally in psychology, researchers make multiple linear regression models with predictor variables, such as the Big Five personality traits, age, and gender, to predict outcomes of interest, such as mental health, as the criterion variable. The models we introduce here are similar. The criterion variable works exactly the same. But instead of personality traits and demographics as the predictor variables, word embeddings are the predictor variables. Compared with personality traits and demographics as predictor variables (seven predictors), word embeddings commonly consist of hundreds or even thousands of dimensions (i.e., referred to here as "predictor variables"). An observant reader understands that such a model is likely to violate the assumption of multicollinearity (i.e., correlated predictor variables). To deal with this, models in the L-BAM Library use slightly more advanced forms of multiple linear regressions (e.g., ridge regression) that reduce the impact of irrelevant predictors through a penalty (a penalty, represented by a number, pushes abundant predictors toward 0, and the higher the penalty is, the stronger it pushes predictors toward 0). Furthermore, in a standard multiple regression, the predictors are fit to the criterion in one single model without testing if this fit works on new data. All models here, instead, have first fitted the criterion using various penalties on the predictors from a portion of the data. They are then tested on the remaining portion of the data for their predictive accuracy, and the degree of penalty is also evaluated. This process, known as "cross-validation," is essential to secure the generalizability of models. The most common way to develop LBAMs is by using ridge regression via the `textTrain()` function in the *text* package (which have been described in detail in Kjell, Giorgi, & Schwartz, 2023). Thus, the `textTrain()` function is one way to examine the relationship between language and a criterion and hence creates an LBAM. However, alternative methods (than ridge regression) also exist, such as fine-tuning large language models. These models can subsequently be uploaded online (e.g., to the OSF or GitHub) and added to the L-BAM Library.

Applying models from the L-BAM Library. Finally, the LBAMs can be applied on new data, a fully automated process that takes the user's new text data as input and provides a predicted score as outcome. This step is achieved with the `textAssess()` function of the *text* package, as we describe in more detail in the tutorial next.

The *textAssess()* function

In the following, we describe how L-BAM Library users can quickly implement LBAMs on their own data in R using the *text* package. The function for doing this is called `textAssess()`,[1] and this function is used to assign language a psychological assessment, such as generating a depression-severity score based on the language. The main parameters of the function include `model_info` for indicating the LBAM and `texts` or `word_embeddings` for passing the (embedded) language to which the model will be applied (see Code Box 1). The output from the `textAssess()` function is returned as either assessment scores (if the outcome is continuous) or classification labels with probability scores (if the outcome is binary).

For an example of how to download and apply an LBAM using the `textAssess()` function, see Code Box 1. Here, a model is downloaded that has been trained using the *text* package to assess depression severity (Gu et al., 2025) and is applied to two example sentences (in `text_to_assess`). The `textAssess()` function downloads the chosen model (indicated by `model_info`), transforms the example data (indicated by `texts`) into word embeddings, and applies the LBAM to these word embeddings to assess depression-severity

Code Box 1. Example on Depression Severity

```
# Example text to access
text_to_assess = c(
  "I feel down and blue all the time.",
  "I feel great and have no worries that
  bother me.")

library(text) # See Code Box 2 if the package has not
been installed.

# Predict depression-severity scores using a model
   trained with the text package.
# Download the model, create word embeddings, and
   apply the model to these word embeddings.
depression_scores <- textAssess(
  model_info = "depression_text_phq9_
    roberta23_gu2024",
  texts = text_to_assess,
  dim_names = FALSE)

# Output
depression_scores

# A tibble: 2 × 1
   `word_embeddings__PHQ9$PHQtotpred`
                                  <dbl>
1                                 17.2
2                                 4.10
```

Code Box 2. Install the *text* Package

```
# Step 1: Install the text package in R.
install.packages("text")
library(text)

# Step 2: Set up the required Python environment.
# The function first checks that common system
  dependencies are satisfied and if not, provides
  instructions for how to install them in the terminal.
textrpp_install()

# Step 3: Initialize the Python environment for use
  with text.
textrpp_initialize()

# If you encounter any issues during installation
  or setup, refer to the latest instructions and
  troubleshooting guide at
https://r-text.org/articles/ext_install_guide.html.
```

Code Box 3. Example on Valence and Well-Being

```
# Assess the valence of the harmony-in-life texts.
# Download the model, create word embeddings, and
  apply the model to these word embeddings.
    valence_scores <- textAssess(
    model_info = "valence_facebook_mxbai23_
      eijsbroek2024",
    texts = Language_based_assessment_
      data_8$satisfactiontexts)

# Correlate the assessed valence scores with the
  harmony-in-life scores.
cor(valence_scores$texts__Valencepred,
   Language_based_assessment_data_
     8$swlstotal)

[1] 0.7421613
```

scores. The output includes the assessed depression-severity scores for the two example responses, 17.2 and 4.10 on the Patient Health Questionnaire–9 (PHQ-9; Kroenke & Spitzer, 2002). These predictions show face validity given that the first response sounds more depressed than the second response.

Code Box 2 shows how to set up the *text* package after installation, which is necessary the first time using the package. Because Python (another coding language) is used at the forefront of most large-language-model development and deployment, the *text* package relies on Python-based tools to access cutting-edge functionality. Under the hood, *text* sets up a dedicated Python environment using Miniconda (a lightweight collection of prebuilt tools for managing Python environments and packages). It then automatically installs key libraries, such as Hugging Face Transformers and PyTorch. This setup enables R users to seamlessly access powerful language models without needing to manually install or configure all the Python dependencies.

Some Python libraries require system-level dependencies that vary across operating systems and platforms. The *text* package automatically checks for these dependencies and if any are missing, provides instructions on how to install them. In some cases, this may require you to download and install tools using the Terminal. More information about platform-specific requirements and troubleshooting is available at https://r-text.org/articles/ext_install_guide.html. To ensure broad compatibility, the installation process and most of the package functionality are automatically and continuously tested on GitHub Actions across macOS, Windows, and Ubuntu systems.

For users who prefer not to install anything locally, we offer the ability to run the tutorial directly in Google Colab, requiring no setup on your own machine. Whenever using this option, make sure to follow the privacy concerns regarding your data.

Code Box 3 shows an example of how to download an LBAM to assess valence and apply it on satisfaction-with-life descriptions using the `textAssess()` function. The valence model was trained using the *text* package to assess human-annotated valence from Facebook posts (Eijsbroek et al., 2026). The satisfaction-with-life descriptions are part of example data of the *text* package (`Language_based_assessment_data_8`) and include text responses in which participants described their experienced satisfaction with life (`satisfactiontexts`) and self-reported ratings of their satisfaction with life (`swlstotal`; Diener et al., 1985). The `textAssess()` function downloads the valence model, transforms the satisfaction descriptions into word embeddings, and applies the downloaded model to these word embeddings to assess valence scores. We correlated the assessed valence scores to the participant's self-reported satisfaction-with-life scores ($r = .74$). This strong positive correlation shows construct validity given that higher valence scores indicate positive emotion, which should theoretically correspond to a higher level of satisfaction with life.

Code Box 4 shows another example of how to download LBAMs for assessing implicit motives and applying it on harmony-in-life descriptions using the `textAssess()` function. The two implicit-motive models were trained to assess the need for affiliation and power from expert-coded stories from picture-story exercises (Nilsson et al., 2024). They were applied to text responses in which participants described their experienced harmony in life (`harmonytexts`), part of the *text* package example data (`Language_based_assessment_data_8`). The `textAssess()` function downloads the implicit-motive models, transforms

Code Box 4. Example of Implicit Motives

```
# Assign implicit-motives labels to the satisfaction
  texts using a fine-tuned model.
# Download the models, create word embeddings, and
  apply the models to these word embeddings.
implicit_affiliation <- textAssess(
   model_info = "implicitaffiliation_
     roberta_ft_nilsson2024",
   texts = Language_based_assessment_
     data_8$harmonytexts)

implicit_power <- textAssess(
 model_info = "implicitpower_roberta_ft_
   nilsson2024",
 texts = Language_based_assessment_
   data_8$harmonytexts)

# Test difference in affiliation versus power in the
  harmony texts.
affiliation <- implicit_affiliation$.pred_1
power <- implicit_power$.pred_1

t.test(affiliation, power)

        Welch Two Sample t-test

data: affiliation and power
t = 4.1368, df = 48.011, p-value = 0.000141
alternative hypothesis: true difference in
   means is not equal to 0
95percent confidence interval:
 0.1447024 0.4183845
sample estimates:
 mean of x mean of y
0.34853536 0.06699195
```

the harmony-in-life descriptions into word embeddings, and applies the implicit-motive model to these word embeddings to assess implicit needs of affiliation and power. We performed a *t* test to test the difference between affiliation and power in the harmony-in-life texts, showing that the implicit need of affiliation (M = .25) was significantly more present than the implicit need of power (M = .07) in texts in which people describe their experienced harmony in life (p < .001). The difference shows high construct validity given that harmony in life is empirically more related to relationships and belonging than power (Kjell et al., 2016; Lomas et al., 2025).

Code Box 5 shows how the L-BAM Library can be examined in R. It is possible to import the library as a data frame using the `textLBAM()` function to easily search for applicable models by filtering the models based on your construct of interest. Here, we show how to filter for the current eight depression models. It is also possible to read individual models and retrieve descriptive information one by one.

Input features: language, word embeddings, and other variables. The *textAssess* function requires predictors in the form of language features and/or other variables (see Table 2). Most models can take either raw language (`texts`) or word embeddings (`word_embeddings`). When language is provided, `textAssess()` will retrieve information from the model object (`model_info`) about the required word embeddings, specifying which large language model to use and its configuration (e.g., the layer or layers to use).

The model and word embeddings will automatically be saved in the working directory when using a model trained with the *text* package. The function first checks if the working directory already has computed word embeddings for a given text; if not, the function retrieves them from the specified large language model (using the `textEmbed()` function; see Kjell, Giorgi, & Schwartz, 2023). However, if you pass word embeddings (with the argument `word_embeddings`) directly, it is crucial to remember that the word embeddings must match those on which the model was trained (i.e., the language is transformed into word embeddings with the same model as indicated in `textAssess()` function).

Furthermore, some models have been trained using additional predictors other than language, that is, more than word embeddings, such as gender and age. In those cases, these variables are appended as a data frame using the x_append parameter. To know whether additional predictors are needed, you can access the model object to see if and what variables are needed for x_append (see information under “x_append” in the `model$model_description`). It is also possible to create and use models based solely on x_append features (i.e., no language features as predictors), which can be useful when comparing an LBAM with a model that uses only demographic variables as predictors while keeping the methods consistent.

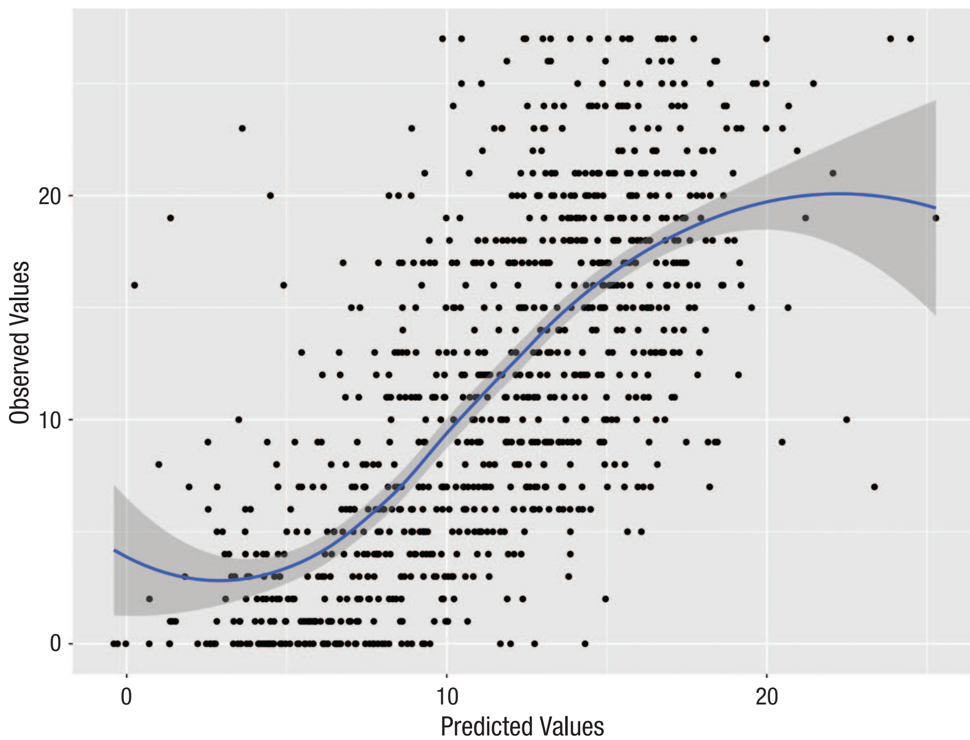
Fine-tuned models. The *textAssess* function by default uses a model object trained in R with the *text* package (called “text-trained”) but can also use fine-tuned models. A text-trained model, which most of the models in the L-BAM Library are, is typically based on a predictive model algorithm (e.g., ridge regression) that has been trained on word embeddings to predict an outcome using a text-train function of the *text* package (e.g., `textTrain()` or `textTrainRegression()`; see Kjell, Giorgi, & Schwartz, 2023). For researchers who want to contribute to the L-BAM Library, these are good functions for model development and are described in detail in Kjell et al. (2023). A “fine-tuned” model is a large language model (e.g., RoBERTa; Y. Liu et al., 2019) that has received further training (i.e., it has been fine-tuned) for a specific task (e.g., classifying a text as positive or negative) or domain (e.g., to model clinical or

Code Box 5. Examine Language-Based Assessment Model (L-BAM) Library

```
# Import the language-based assessment model as a data frame and filter for models starting with "dep."
textLBAM(
   columns = c("Construct_Concept_Behaviours", "Name"),
   construct_start = "dep",
   lbam_update = FALSE
)

  Construct_Concept_Behaviours                                                   Name
1                   depression                  depression_select_phq9_roberta23_gu2024
2                   depression                   depression_words_phq9_roberta23_gu2024
3                   depression                 depression_phrases_phq9_roberta23_gu2024
4                   depression                    depression_text_phq9_roberta23_gu2024
5                   depression                  depression_select_cesd_roberta23_gu2024
6                   depression                    depression_word_cesd_roberta23_gu2024
7                   depression                  depression_phrase_cesd_roberta23_gu2024
8                   depression                    depression_text_cesd_roberta23_gu2024

# Retrieve information about a trained model with the text package in R.
model_Gu2024 <- readRDS("depressiontext_robertaL23_phq9_Gu2024.rds")

# To see the model's training performance, run the following:
model_Gu2024$results

      Pearson's product-moment correlation

data: predy_y$predictions and predy_y$y
t = 29.765, df = 954, p-value < 2.2e-16
alternative hypothesis: true correlation is greater than 0
95percent confidence interval:
 0.6652463 1.0000000
sample estimates:
      cor
0.6939053

# The model typically also includes the target with cross-validated predictions.
> model_Gu2024$predictions
# A tibble: 963 × 3
   predictions     y id_nr
         <dbl> <dbl> <int>
 1        6.91     0     1
 2        8.52    17     2
 3       13.4     21     3
 4       15.8     13     4
 5       14.1      9     5
 6        2.82     1     6
 7       11.0      9     7
 8       11.3      5     8
 9        3.61    23     9
10        8.32     2    10
# ℹ 953 more rows
# ℹ Use `print(n = . . .)` to see more rows

# These can be used to examine the performance further.
ggplot(model_Gu2024$predictions, aes(x = predictions, y = y)) +
  geom_point() +
  geom_smooth(method = "loess") +
  labs(
    x = "Predicted Values",
    y = "Observed Values",
   caption = "Cross-validated predicted vs. observed PHQ-9 scores (Gu et al., 2024)."
  )
```

(continued)

Code Box 5. *(continued)*



social media language more accurately). Thus, the model parameters of the large language model have been adjusted for the new task or domain. Most fine-tuned models in the L-BAM Library take only language (`texts`) as input (i.e., no `word_embeddings` and no additional predictors from x_append). Fine-tuned models can be found at the repository `huggingface.co` and can be developed in the *text* package with the functions *`textFineTuneTask()`* and *`textFineTuneDomain()`*. For more information about fine-tuning models, see Demszky et al. (2023).

Table 2. Main Function Parameters and Arguments for the *textAssess()* Function

Parameter	Description
model_info	A pretrained model object or a path (string) to a model, which can be a path to a model stored locally, a name from the Language-Based Assessment Model Library, or an online path
texts	A character string or a data frame with a character column to assess
word_embeddings	Word embeddings (works only for text-trained models and cannot be combined with texts)
x_append	A data frame with additional variables to use for prediction

The L-BAM Library: reproducibility, replication, and generalizability

To make model sharing more straightforward and accessible, we introduce the L-BAM Library. The L-BAM Library is a searchable online database (https://r-text.org/articles/LBAM.html[2]) in which users can search for models and filter according to different model characteristics, such as the type of construct, the model's predictive accuracy, or the type of language.

The L-BAM Library aims to comprise the most relevant information for model sharing, balancing thoroughness with practicality. We expect that most well-validated models will be accompanied by additional documentation, such as a peer-reviewed article, a model card describing the model in depth (Mitchell et al., 2019), and/or a data sheet clearly describing the training data set (Gebru et al., 2021). We also encourage reporting guidelines, such as the TRIPOD(+AI) statement for transparent reporting of research that develops, validates, or extends (updates) prediction models (Collins et al., 2024) and the LEADING statement for comprehensive reporting of how best-estimate assessments are achieved for training/evaluating prediction models that assess (psychiatric or medical) conditions lacking a more objective truth or "gold standard" (Eijsbroek et al., 2025).

Table 3. Overview of the Language-Based Assessment Model Library

Information type	Description and categories	Example[a]
Outcome (five categories)	Information about the criterion variable of the model, such as - Construct/behaviors - Outcome - Language - Language type - Level of analysis	Nine-item Patient Health Questionnaire (depression rating)
Training data (eight categories)	Information about the language data trained to assess the outcome, such as - *N* training - *N* evaluation - Data source - Participant information - Whether training data are open - Language prompt (if applicable)	Descriptions of depression from the prompt "Have you been depressed in the past two weeks?" 967 participants recruited through Prolific
Model type (two categories)	Information about the algorithm that fitted the word embeddings of the training data to the outcome, such as - Model type - Features	Ridge regression with Bert Base Layer 11
Model performance (nine categories)	The performance of the model's predictions, such as - Validation metric - Cross-validated accuracy - Held-out accuracy - Model-preregistration accuracy - Other evaluations	Pearson r = .70 in held-out accuracy
Ethical considerations (two categories)	Ethical information, including - Ethical approval - Ethical statement	The Swedish National Ethics Review Board deemed this research study exempt from requiring ethical approval.
Metadata (eight categories)	Useful information about the model, such as - Study type (e.g., development or usage) - Reference - Model-creation date - Contact details - License	Gu et al. (2024), open
Access (four categories)	Information on how to assess the model, such as - Name to retrieve with the *textAssess()* function - Name description - Path	depression_words_phq9_roberta23_gu2024

[a]The content of the categories is detailed in Tables S1 through S3 in the Supplemental Material available online, including examples of them.

Using and contributing to the L-BAM Library

Next, we outline the key information types of the L-BAM Library, offering a standardized format for describing the models, including outlining aspects of the outcome and training data, model performance and ethical considerations, and metadata and access (Table 3). These components are relevant for users to understand the models they use and for contributors who should report these when adding models to the L-BAM Library. Note that they are described briefly here and in more detail in the L-BAM Library (see https://docs.google.com/spreadsheets/d/14PcfTwQJZCKbSh6ylOq1Qm1VT44X4RD0aR4-dt6bink/edit?gid=194707973#gid=194707973) and in Tables S1 through S3 in the Supplemental Material.

The outcome section describes what the model predicts, assesses, or classifies, such as the psychological construct or behavior (e.g., depression through PHQ-9), and details about the specific outcome the model was trained to and the type of language used to train the model (see Table S1 in the Supplemental Material).

The training-data section details the data set used to train the model, including the number of observations, where the data are attained (e.g., online, clinic), participants (e.g., demographics), and the type of labels used in training (e.g., self-reported). It also describes whether the model includes information about the language distribution (a word-frequency table) used in training to assess language similarity with new data (see Table S1 in the Supplemental Material).

The model section focuses on the technical aspects of the model, including the type of model used for prediction (e.g., ridge regression) and the features used for prediction, such as word embeddings and/or demographic information (see Table S2 in the Supplemental Material).

The model-performance section presents the key performance metrics of the model; one can include one primary metric (e.g., Pearson *r* or area under the curve), which is possible to filter, and then give additional relevant validation metrics (e.g., mean absolute error, sensitivity). One can also include accuracy from several evaluation frameworks, including (nested) cross-validation, held-out accuracy, and Sequential Evaluation With Model Pre-registration (SEMP; Kjell, Ganesan, et al., 2024; see Table S2 in the Supplemental Material). SEMP aims to address concerns that predictive models often underperform in independent or prospective samples (e.g., Chekroud et al., 2024; Kernbach & Staartjes, 2022; Spasic & Nenadic, 2020; also see Essential Caveat About Generalization section). It essentially involves preregistering LBAMs and expected outcomes before applying them to held-out evaluation data. If the results replicate with similar effect sizes, this adds strong evidence for the model's robustness and generalizability. Note that not all models will include these estimates because they depend on how the model was developed and evaluated.

The ethical-considerations section includes the ethical-approval application ID associated with the model's development (if applicable) and outlines ethical considerations or concerns addressed during the development and testing phases and those to consider in future applications (see Table S2 in the Supplemental Material).

The model-metadata-and-access section describes the study type (e.g., development or replication), citation details, licensing restrictions, and where the model can be accessed (see Table S3 in the Supplemental Material). If specific commands for using the models are applicable, they should also be mentioned here. We once again stress that all the information of what to add in the documentation is described in the L-BAM Library at https://r-text.org/articles/LBAM.html. Furthermore, we have uploaded a template with the headings of the library so that researchers can fill everything in on their local computer before adding all information to the library itself.

The L-BAM Library versus other repositories

The L-BAM Library focuses on social sciences and in particular, psychology, offering a standardized and accessible collection of models that are fully compatible with R and can be easily applied using `textAssess()`. By streamlining the application process, the L-BAM Library minimizes technical barriers, making it easier for researchers to integrate LBAs into their work. The L-BAM Library is not a repository—it is a library from which the actual models are hosted on repository platforms, such as GitHub, Hugging Face, and OSF. What further sets the L-BAM Library apart is its structured framework specifically designed for LBAs, in which models are accompanied by comprehensive documentation to help researchers quickly gain an overview of available models and identify relevant models for their research needs. Thus, the L-BAM Library promotes open and transparent scientific practices, emphasizing reproducibility, independent validation, and accessibility. The library is open for anyone to contribute models, meaning that it does not curate or filter models itself but instead provides a structured system for sharing and applying them in research. This openness encourages collaboration, refinement, and broader validation and application of language-based psychological assessments.

Ethical considerations, responsibility, and AI safety

The scores from LBAs can be used for further statistical analyses, such as standard hypothesis testing or predictive modeling. However, using LBAs comes with several ethical considerations.

Privacy. Language data are typically highly informative and personal, making it hard to anonymize, and researchers must consider ethical challenges and privacy issues in all steps of data collection, storage, and analyses of natural language (see e.g., Leidner & Plachouras, 2017). The L-BAM Library includes models that can be downloaded and run locally in the user's own environment, allowing the user to avoid sharing sensitive information with a third party (e.g., ChatGPT). However, when uploading models, it is crucial to remember that certain types, such as fine-tuned large language models and text-trained models with language-distribution data, may contain sensitive information; hence, it is crucial to consider privacy concerns before sharing models. A text-trained model without a saved language distribution contains no language data on which it was trained.

Responsible applications and generalizability. The L-BAM Library does not involve peer review of models or warranty for the models it includes. As a class of techniques, research has shown that LBAs have comparable or

Box 1. Responsible Sharing and Use of Language-Based Assessment Models (LBAMs)

Guidelines for contributors

To support transparency, traceability, and responsible use, contributors should do the following:

1. **Fill out the LBAM submission sheet:** Fill out the model-submission form available at https://r-text.org/articles/LBAM.html. This includes key metadata such as model name, outcome variable(s), model type, training-data size, validation metrics, ethical considerations, and relevant metadata (e.g., links to articles or preprints).
2. **Public hosting:** Host the model on a publicly accessible platform with version control—such as OSF, Hugging Face, GitHub, or Bitbucket—to ensure reproducibility and long-term access.
3. **Contact:** Email a library maintainer (see contact details at https://r-text.org/articles/LBAM.html) using the same email address that you provide under contact_details in the metadata.

Once these steps are completed, we will publish the model in the Language-Based Assessment Model (L-BAM) Library, making it accessible to the broader research community.

Guidelines for users

Before using a model from the L-BAM Library, we recommend the following steps to ensure its suitability for your research context:

1. **Verify source and contact information**
   Review the listed contact details. If in doubt, reach out to the model contributor for clarification. Using models may carry security risks, including the possibility of malicious code in the file. Each model in the L-BAM Library must include a designated contact person with a valid email address (i.e., not placeholder or fabricated information) and must be hosted on publicly accessible platforms with version control, such as OSF, to ensure transparency and traceability. Always review and make sure you trust the source of any model you load. In situations requiring extra caution, consider loading models in secure, isolated environments rather than directly into your local R session (e.g., see Docker). Check for linked preprints or peer-reviewed publications that describe the model.
2. **Critically evaluate the development and validation details**
   Examine how the model has been validated, including reported performance metrics (e.g., root mean square error, correlations) and the populations used for validation. Ensure these align with your intended use case.
3. **Test generalizability**
   Evaluate the model's performance in your specific context by applying it to some new data that include the outcome measure you are targeting (e.g., if you have 10,000 texts you want to apply the LBAM on, make sure you have the criterion variable associated with at least a subset of the 10,000 texts, such as 100). This allows for direct testing of generalizability.

exceeding validity and reliability as traditional rating scales (Kjell, Kjell, & Schwartz, 2024). However, each specific instance of such an assessment must undergo rigorous evaluation for validity and reliability for target populations and use contexts before being trusted (just as any rating scale would). Although information about the models and how to access them is provided in the L-BAM Library, it is essential for users to independently assess the accuracy, suitability, validity, and reliability of each model for their specific research needs (for details about responsible sharing and usage of LBAMs, see Box 1).

Importantly, evaluating the suitability of a model includes explicit evaluation of generalizability. Our proposed "gold standard" for testing generalizability is to have similar assessments as the model was trained on in a subset of the new data. For example, suppose an LBAM has been trained to assess depression ratings from clinical interviews and a researcher wants to assess depression severity from social media language. In that case, the researcher should make sure there are enough participants having both social media language and depression-severity scores that the model's generalizability can be tested on. The required sample size should be determined by the desired precision in estimating the model's accuracy in the new sample, with attention to the width of its confidence interval (e.g., the confidence interval around a correlation coefficient). We propose this procedure as the "gold standard" because it is testing the model on a subset of the data it will subsequently be applied on.

When such paired data are unavailable, researchers may instead explore differences or similarities in language distributions between the training and target data sets. In the Supplemental Material, we show that one such approach—calculating target recall between training and test data—correlates meaningfully with generalizability performance (rs = .38–.39; n = 68 tests; see Table Box 1 in the Supplemental Material). This suggests target recall may offer a useful proxy for estimating generalizability, although further research is needed to refine and validate such distributional metrics.

Ultimately, the long-term goal is to accumulate enough well-documented model evaluations across diverse settings to enable meta-analyses and potentially predictive benchmarks of generalizability in new language contexts. This requires community-wide participation and is a key direction for future development of the LBAM ecosystem.

Ethical principles. Finally, there are several ethical principles (Jobin et al., 2019; Peters et al., 2020), regulations, and legal frameworks (European Commission, 2023; Hauglid & Mahler, 2023; U.S. Food & Drug Administration, 2021; Veale & Zuiderveen Borgesius, 2021; White House Office of Science and Technology Policy, 2022) concerning the development and use of AI and large language models. A review of more than 80 international guidelines identified five key ethical principles: transparency, justice and fairness, nonmaleficence, responsibility, and privacy (Jobin et al., 2019). We encourage users to explore, apply, and stay current with these resources.

Summary

In this tutorial, we presented the `textAssess()` function that enables researchers to get scores on psychological constructs from language by using preexisting models from an open library of LBAMs (the L-BAM Library). The library aims to assist researchers in analyzing language data while supporting replication, independent validation, and broader model application. Doing so is expected to promote resource efficiency and enhance comparability when models are applied across different research groups and studies. After reading this tutorial, researchers should possess the skills to apply any model from the L-BAM Library to their language data (for recommended further readings, see Table 4).

We encourage researchers to expand the library with models predicting both psychological constructs and other social-science outcomes. We hope the library will help researchers to (a) use previously developed models, (b) upload information about their models, and (c) report relevant research that further validates existing models. By encouraging independent validation and transparency, the L-BAM Library can hopefully help strengthen research rigor on LBAs and advance the field.

Table 4. Recommended Reading

Collins et al. (2024), "TRIPOD+ AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods," *The BMJ*.
Gebru et al. (2021), "Datasheets for Datasets." *Communications of the ACM*.
Kjell et al. (2023), "The text-Package: An R-Package for Analyzing and Visualizing Human Language Using Natural Language Processing and Transformers," *Psychological Methods*.
Mitchell et al. (2019), "Model Cards for Model Reporting," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*.

Transparency

Action Editor: David A. Sbarra
Editor: David A. Sbarra
Author Contributions
August H. Nilsson: Conceptualization; Formal analysis; Investigation; Methodology; Writing – original draft; Writing – review & editing.
Veerle C. Eijsbroek: Formal analysis; Software; Validation; Writing – review & editing.
Zhuojun Gu: Data curation; Formal analysis; Methodology; Resources; Software.
Katarina Kjell: Data curation; Resources; Writing – review & editing.
Salvatore Giorgi: Resources; Software; Writing – review & editing.
Roman Kotov: Validation; Writing – review & editing.
Adithya V. Ganesan: Data curation; Formal analysis; Resources; Software.
H. Andrew Schwartz: Conceptualization; Data curation; Formal analysis; Resources; Software; Supervision; Writing – review & editing.
Oscar N. E. Kjell: Conceptualization; Data curation; Formal analysis; Funding acquisition; Methodology; Resources; Software; Supervision; Validation; Writing – original draft; Writing – review & editing.
Declaration of Conflicting Interests
K. Kjell and O. Kjell have cofounded a start-up using computational language assessments to assess mental-health problems.
Funding
V. C. Eijsbroek, K. Kjell, and O. N. E. Kjell were supported by Forskningsrådet för hälsa, arbetsliv och välfärd (FORTE) (STY-2022/0007; 2022-01022); O. N. E. Kjell was also funded by Marianne och Marcus Wallenbergs stiftelse (MMW 2021.0058). Z. Gu was funded by a grant from e-Science Collaboration at Lund University (eSSENCE@LU) (7:5). Computations were enabled by resources provided by the Swedish National Infrastructure for Computing at Chalmers

University of Technology, partially funded by the Swedish Research Council through Grant Agreement No. 2018-05973. This work was also supported in part by a grant from the Centers for Disease Control and Prevention / National Institute for Occupational Safety and Health (CDC/NIOSH) (U01 OH012476) and a grant from the National Institutes of Health-National Institute on Alcohol Abuse and Alcoholism (NIH-NIAAA) (R01 AA028032).

Open Practices

This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iDs

August H. Nilsson https://orcid.org/0000-0003-2416-3755
Veerle C. Eijsbroek https://orcid.org/0009-0000-7632-4085
Zhuojun Gu https://orcid.org/0009-0000-1610-4830
Roman Kotov https://orcid.org/0000-0001-9569-8381
Adithya V. Ganesan https://orcid.org/0000-0001-6179-624X
Oscar N. E. Kjell https://orcid.org/0000-0002-2728-6278

Acknowledgments

A preprint of this manuscript is available on PsyArXiv. Open code and data available at https://r-text.org/, https://cloud.r-project.org/web/packages/text/index.html, https://github.com/OscarKjell/text/, and https://osf.io/q97s2/.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/25152459261419036.

Notes

1. Interchangeable aliases of *textAssess()* are *textPredict()* and *textClassify()*, which work the same but are named differently to align with different tasks and contexts theoretically. For example, depending on the specific model, the functions can be used to predict future mental states (e.g., an estimate of future depression severity), assess a current mental state (e.g., a harmony-in-life score), or classify texts by assigning labels representing personal characteristics (e.g., need for power, achievement, or affiliation).
2. The table is developed using the R package *reactable* (Lin, 2023).

References

Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, *6*(4), 529–542.

Argamon, S., Koppel, M., Pennebaker, J. W., & Schler, J. (2007). Mining the blogosphere: Age, gender and the varieties of self-expression. *First Monday*, *12*(9). https://doi.org/10.5210/fm.v12i9.2003

Bayram, A. B., & Ta, V. P. (2019). Diplomatic chameleons: Language style matching and agreement in international diplomatic negotiations. *Negotiation and Conflict Management Research*, *12*(1), 23–40. https://doi.org/10.1111/ncmr.12142

Boyd, R. L., Morrison, N. R., Horwitz, S. D., Maciag, R., Travers-Hill, E., & Kim, Y. (2024). Are we listening to every word? Using multiple analytic methods to examine qualitative data. *Cogent Mental Health*, *3*(1), 2433791.

Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, *40*(1), 21–41.

Brede, M., Schönbrodt, F., Hagemeyer, B., & Lerche, V. (2025, May). Automatically coding implicit motives in picture story exercises: The automated motive coder. In *Proceedings of the ICWSM Workshops (Workshop on Data for the Well-being of Most Vulnerable, Copenhagen, Denmark, June 23)*. Association for the Advancement of Artificial Intelligence (AAAI). https://workshop-proceedings.icwsm.org/pdf/2025_31

Bucur, A. M., Podină, I. R., & Dinu, L. P. (2021). *A psychologically informed part-of- speech analysis of depression in social media*. arXiv. https://doi.org/10.48550/arXiv.2108.00279

Chekroud, A. M., Hawrilenko, M., Loho, H., Bondar, J., Gueorguieva, R., Hasan, A., Kambeitz, J., Corlett, P. R., Koutsouleris, N., Krumholz, H. M., Krystal, J. H., & Paulus, M. (2024). Illusory generalizability of clinical prediction models. *Science*, *383*(6679), 164–167. https://doi.org/10.1126/science.adg8538

Chen, J., Qiu, L., & Ho, M. H. R. (2020). A meta-analysis of linguistic markers of extraversion: Positive emotion and social process words. *Journal of Research in Personality*, *89*, Article 104035. https://doi.org/10.1016/j.jrp.2020.104035

Chen, Y., Li, J., Song, C., Zhao, Q., Tong, Y., & Fu, G. (2024). *Deep learning and large language models for audio and text analysis in predicting suicidal acts in Chinese psychological support hotlines*. arXiv. https://doi.org/10.48550/arXiv.2409.06164

Chollet, F. (2015). *Keras*. https://keras.io

Collins, G. S., Moons, K. G., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., Ghassemi, M., Liu, X., Reitsma, J. B., & Van Smeden, M. (2024). TRIPOD+ AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *The BMJ*, *385*, Article e078378. https://doi.org/10.1136/bmj-2023-078378

Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying mental health signals in Twitter. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (pp. 51–60). Association for Computational Linguistics. https://doi.org/10.3115/v1/W14-3207

Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., Eichstaedt, J. C., Hecht, C., Jamieson, J., Johnson, M., Jones, M., Krettek-Cobb, D., Lai, L., Jones Mitchell, N., Ong, D. C., Dweck, C. S., Gross, J. J., & Pennebaker, J. W. (2023). Using large language models in psychology. *Nature Reviews Psychology*, *2*, 688–701. https://doi.org/10.1038/s44159-023-00241-5

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies volume 1 (long and short papers)* (pp. 4171–4186). Association for Computational Linguistics.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75.

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., Jha, S., Agrawal, M., Dziurzynski, L. A., & Sap, M. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological Science*, *26*(2), 159–169.

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoțiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences*, *115*(44), 11203–11208.

Eijsbroek, V. C., Kjell, K., Schwartz, H. A., Boehnke, J. R., Fried, E. I., Klein, D. N., . . . Kjell, O. N. (2025). The LEADING guideline: Reporting standards for expert panel, best-estimate diagnosis, and longitudinal expert all data (LEAD) methods. *Comprehensive Psychiatry*, *141*, 152603. https://doi.org/10.1016/j.comppsych.2025.152603

Eijsbroek, V. C., Nilsson, A., Ackermann, L., Gu, Z., Wiebel, C., Ganesan, A. V., . . . Kjell, O. N. E. (2026, March 13). *Multiple methods for visualizing human language: A tutorial for social and behavioural scientists*. PsyArXiv. https://doi.org/10.31234/osf.io/nxfvr_v2

European Commission. (2023). *CE marking*. https://single-market-economy.ec.europa.eu/single-market/ce-marking_en

Fausey, C. M., & Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, *17*(5), 644–650.

Ganesan, A. V., Matero, M., Ravula, A. R., Vu, H., & Schwartz, H. A. (2021). Empirical evaluation of pre-trained transformers for human-level NLP: The role of sample size and dimensionality. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4515–4532). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.357

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., III, & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86–92. https://doi.org/10.1145/3458723

Gu, Z., Eijsbroek, V. C., Kjell, K., Wiebel, C., Järvholm, K., Schwartz, H. A., & Kjell, O. N. E. (in progress). *Understanding suicidality and self-harm through probed open-ended language: A sequential evaluation with model pre-registration*.

Gu, Z., Kjell, K., Schwartz, H. A., & Kjell, O. (2024). *Natural language response formats for assessing depression and worry with large language models: A sequential evaluation with model pre-registration*. PsyArXiv. https://doi.org/10.31234/osf.io/p67db

Gu, Z., Kjell, K., Schwartz, H. A., & Kjell, O. (2025). *Natural language response formats for assessing depression and worry with large language models: A sequential evaluation with model pre-registration*. Assessment, https://doi.org/10.1177/10731911251364022

Hauglid, M. K., & Mahler, T. (2023). Doctor Chatbot: The EU's regulatory prescription for generative medical AI. *Oslo Law Review*, *10*(1), 1–23. https://doi.org/10.18261/olr.10.1.1

Hu, T., Xiao, H., Luo, J., & Nguyen, T. V. T. (2016). What the language you tweet says about your occupation. *Proceedings of the International AAAI Conference on Web and Social Media*, *10*(1), 181–190.

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, *22*(1), 39–44.

Jaidka, K., Giorgi, S., Schwartz, H. A., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2020). Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences*, *117*(19), 10165–10171.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.

Jose, R., Matero, M., Sherman, G., Curtis, B., Giorgi, S., Schwartz, H. A., & Ungar, L. H. (2022). Using Facebook language to predict and describe excessive alcohol use. *Alcoholism: Clinical and Experimental Research*, *46*(5), 836–847. https://doi.org/10.1111/acer.14807

Kernbach, J. M., & Staartjes, V. E. (2022). Foundations of machine learning-based clinical prediction modeling: Part II—Generalization and overfitting. In V. E. Staartjes, L. Regli, & C. Serra (Eds.), *Machine learning in clinical neuroscience* (Vol. 134, pp. 15–21). Springer International Publishing. https://doi.org/10.1007/978-3-030-85292-4_3

Kjell, O., Daukantaitė, D., Hefferon, K., & Sikström, S. (2016). The harmony in life scale complements the satisfaction with life scale: Expanding the conceptualization of the cognitive component of subjective well-being. *Social Indicators Research*, *126*(2), 893–919. https://doi.org/10.1007/s11205-015-0903-z

Kjell, O., Daukantaitė, D., & Sikström, S. (2021). Computational language assessments of harmony in life—not satisfaction with life or rating scales—correlate with cooperative behaviors. *Frontiers in Psychology*, *12*, Article 601679. https://doi.org/10.3389/fpsyg.2021.601679

Kjell, O., Ganesan, A. V., Boyd, R., Oltmanns, J. R., Rivero, A., Feltman, S., Carr, M. A., Luft, B., Kotov, R., & Schwartz, H. A. (2024). *Demonstrating high validity of a new AI-language assessment of PTSD: A sequential evaluation with model pre-registration*. PsyArXiv. https://doi.org/10.31234/osf.io/xw24e

Kjell, O., Giorgi, S., & Schwartz, H. A. (2023). The text-package: An R-package for analyzing and visualizing human language using natural language processing and transformers. *Psychological Methods*, *28*(6), 1478–1498. https://doi.org/10.1037/met0000542

Kjell, O., Giorgi, S., Schwartz, H. A., & Eichstaedt, J. C. (2023). Towards well-being measurement with social media across space, time and cultures: Three generations of progress.

In *World happiness report* (pp. 131–162). Sustainable Development Solutions Network.

Kjell, O., Kjell, K., Garcia, D., & Sikström, S. (2019). Semantic measures: Using natural language processing to measure, differentiate, and describe psychological constructs. *Psychological Methods*, *24*(1), 92–115. https://doi.org/10.1037/met0000191

Kjell, O., Kjell, K., & Schwartz, H. A. (2024). Beyond rating scales: With targeted evaluation, language models are poised for psychological assessment. *Psychiatry Research*, *333*, Article 115667. https://doi.org/10.1016/j.psychres.2023.115667

Kjell, O., Sikström, S., Kjell, K., & Schwartz, H. A. (2022). Natural language analyzed with AI-based transformers predict traditional subjective well-being measures approaching the theoretical upper limits in accuracy. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-07520-w

Kjell, O. N. E., Feltman, S., Schwartz, H. A., Ganesan, A. V., Ringwald, W. R., Clouston, S., . . . Kotov, R. (in progress). *Distinguishing the Language of Mental and Physical Health: A Sequential Evaluation with Model Preregistration of Automated Clinical Visit Interviews*. Retrieved from osf.io/preprints/psyarxiv/wx6ca_v1

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, *32*(9), 1–7.

Kwantes, P. J., Derbentseva, N., Lam, Q., Vartanian, O., & Marmurek, H. H. C. (2016). Assessing the Big Five personality traits with latent semantic analysis. *Personality and Individual Differences*, *102*, 229–233. https://doi.org/10.1016/j.paid.2016.07.010.

Lalk, C., Steinbrenner, T., Kania, W., Popko, A., Wester, R., Schaffrath, J., Eberhardt, S., Schwartz, B., Lutz, W., & Rubel, J. (2024). Measuring alliance and symptom severity in psychotherapy transcripts using Bert topic modeling. *Administration and Policy in Mental Health and Mental Health Services Research*, *51*(4), 509–524. https://doi.org/10.1007/s10488-024-01356-4

Leidner, J. L., & Plachouras, V. (2017). Ethical by design: Ethics best practices for natural language processing. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing* (pp. 30–40). Association for Computational Linguistics.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 1–55.

Lin, G. (2023). *reactable: Interactive data tables for R (R package version 0.4.4)*. https://CRAN.R-project.org/package=reactable

Liu, T., Ungar, L. H., Curtis, B., Sherman, G., Yadeta, K., Tay, L., Eichstaedt, J. C., & Guntuku, S. C. (2022). Head versus heart: Social media reveals differential language of loneliness from depression. *Npj Mental Health Research*, *1*(1), Article 16. https://doi.org/10.1038/s44184-022-00014-7

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *Roberta: A robustly optimized BERT pretraining approach*. arXiv https://doi.org/10.48550/arXiv.1907.11692

Liu, Y., Zhang, X. F., Wegsman, D., Beauchamp, N., & Wang, L. (2022). *POLITICS: Pretraining with same-story article comparison for ideology prediction and stance detection*. arXiv. https://doi.org/10.48550/arXiv.2205.00619

Lomas, T., Nilsson, A. H., Kjell, O., Niemiec, R., Pawelski, J. O., Padgett, R. N., & VanderWeele, T. J. (2025). Differentiating balance and harmony through natural language analysis: A cross-national exploration of two understudied wellbeing-related concepts. *The Journal of Positive Psychology*, *21*(1), 173–191. https://doi.org/10.1080/17439760.2025.2459400

Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, *317*(5834), Article 82. https://doi.org/10.1126/science.1139940

Mesquiti, S., Cosme, D., Nook, E. C., Falk, E. B., & Burns, S. M. (2025). *Predicting psychological and subjective well-being through language-based assessment*. PsyArXiv. https://doi.org/10.31234/osf.io/rfq8p_v1

Mihalcea, R., Biester, L., Boyd, R. L., Jin, Z., Perez-Rosas, V., Wilson, S., & Pennebaker, J. W. (2024). How developments in natural language processing help us in understanding human behaviour. *Nature Human Behaviour*, *8*(10), 1877–1889.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287596

Nilsson, A., Boyd, R., Ganesan, A. V., Kjell, O. N. E., Mahwish, S., Huang, H., Rosenthal, R. N., Ungar, L., & Schwartz, H. A. (2024, November 18). *Language-based assessments for experienced well-being: Accuracy and external validity across behaviors, traits, and states*. PsyArXiv. https://doi.org/10.31234/osf.io/dgnaf

Nilsson, A. H., Hellryd, E., & Kjell, O. (2022). Doing well-being: Self-reported activities are related to subjective well-being. *PLOS ONE*, *17*(6), Article e0270503. https://doi.org/10.1371/journal.pone.0270503

Nilsson, A. H., Runge, J. M., Ganesan, A. V., Lövenstierne, C. V. N. G., Soni, N., & Kjell, O. N. E. (2025). Automatic implicit motives codings are at least as accurate as humans' and 99% faster. *Journal of Personality and Social Psychology*, *128*(6), 1371–1392. https://doi.org/10.31234/osf.io/7s6jp

Nilsson, A. H., Schwartz, H. A., Rosenthal, R. N., McKay, J. R., Vu, H., Cho, Y.-M., Mahwish, S., Ganesan, A. V., & Ungar, L. (2024). Language-based EMA assessments help understand problematic alcohol consumption. *PLOS ONE*, *19*(3), Article e0298300. https://doi.org/10.1371/journal.pone.0298300

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, *108*(6), 934–952. https://doi.org/10.1037/pspp0000020

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., . . . Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural*

Information Processing Systems, *32*, 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*, 547–577.

Perlis, R. H., Goldberg, J. F., Ostacher, M. J., & Schneck, C. D. (2024). Clinical decision support for bipolar depression using large language models. *Neuropsychopharmacology*, *49*(9), 1412–1416. https://doi.org/10.1038/s41386-024-01841-2

Peters, D., Vold, K., Robinson, D., & Calvo, R. A. (2020). Responsible AI—Two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, *1*(1), 34–47.

Sametoğlu, S., Pelt, D. H. M., Eichstaedt, J. C., Ungar, L. H., & Bartels, M. (2024). The value of social media language for the assessment of wellbeing: A systematic review and meta-analysis. *The Journal of Positive Psychology*, *19*(3), 471–489. https://doi.org/10.1080/17439760.2023.2218341

Sarwar, R., Teh, P. S., Sabah, F., Nawaz, R., Hameed, I. A., & Hassan, M. U. (2024). AGI-P: A gender identification framework for authorship analysis using customized fine-tuning of multilingual language model. *IEEE Access*, *12*, 15399–15409. https://doi.org/10.1109/ACCESS.2024.3358199

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., . . . Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS One*, *8*(9), e73791.

Schwartz, H. A., Giorgi, S., Sap, M., Crutchley, P., Ungar, L., & Eichstaedt, J. (2017). DLATK: Differential Language Analysis ToolKit. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 55–60). Association for Computational Linguistics. https://aclanthology.org/D17-2010

Shah, S. V. (2024). Accuracy, consistency, and hallucination of large language models when analyzing unstructured clinical notes in electronic medical records. *JAMA Network Open*, *7*(8), Article e2425953. https://doi.org/10.1001/jamanetworkopen.2024.25953

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Son, Y., Clouston, S. A., Kotov, R., Eichstaedt, J. C., Bromet, E. J., Luft, B. J., & Schwartz, H. A. (2020). *World Trade Center responders in their own words: Predicting PTSD symptom trajectories with AI-based language analyses of interviews*. arXiv. https://doi.org/10.48550/arXiv.2011.064

Spasic, I., & Nenadic, G. (2020). Clinical text data in machine learning: Systematic review. *JMIR Medical Informatics*, *8*(3), Article e17984. https://doi.org/10.2196/17984

Sripada, C., & Taxali, A. (2020). Structure in the stream of consciousness: Evidence from a verbalized thought protocol and automated text analytic methods. *Consciousness and Cognition*, *85*, 103007.

Stade, E. C., Ungar, L., Eichstaedt, J. C., Sherman, G., & Ruscio, A. M. (2023). Depression and anxiety have distinct and overlapping language patterns: Results from a clinical interview. *Journal of Psychopathology and Clinical Science*, *132*(8), 972–983. https://doi.org/10.1037/abn0000850

Sterling, J., Jost, J. T., & Bonneau, R. (2020). Political psycholinguistics: A comprehensive analysis of the language habits of liberal and conservative social media users. *Journal of Personality and Social Psychology*, *118*(4), 805–834. https://doi.org/10.1037/pspp0000275

Sun, J., Schwartz, H. A., Son, Y., Kern, M. L., & Vazire, S. (2020). The language of well-being: Tracking fluctuations in emotion experience through everyday speech. *Journal of Personality and Social Psychology*, *118*(2), 364–387. https://doi.org/10.1037/pspp0000244

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54.

Teferra, B. G., & Rose, J. (2023). Predicting generalized anxiety disorder from impromptu speech transcripts using context-aware transformer-based neural networks: Model evaluation study. *JMIR Mental Health*, *10*(1), Article e44325. https://doi.org/10.2196/44325

Tidwell, C. A., Danvers, A. F., Pfeifer, V. A., Abel, D. B., Alisic, E., Beer, A., Bierstetel, S. J., Bollich-Ziegler, K. L., Bruni, M., Calabrese, W. R., Chiarello, C., Demiray, B., Dimidjian, S., Fingerman, K. L., Haas, M., Kaplan, D. M., Kim, Y. K., Knezevic, G., Lazarevic, L. B., . . . Mehl, M. R. (2025). Are women really (not) more talkative than men? A registered report of binary gender similarities/differences in daily word use. *Journal of Personality and Social Psychology*, *128*(2), 367–391. https://doi.org/10.1037/pspp0000534

U.S. Food & Drug Administration. (2021). *Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan* (Technical Report No. 1).

Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*. ArXiv. https://doi.org/10.48550/arXiv.1706.03762

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97–112.

White House Office of Science and Technology Policy. (2022). *Blueprint for an AI bill of rights making automated systems work for the American people*. https://www.govinfo.gov/content/pkg/GOVPUB-PREX23-PURL-gpo193638/pdf/GOVPUB-PREX23-PURL-gpo193638.pdf

Wiebel, C., Eijsbroek, V. C., Varadarajan, V., Kjell, K., Schwartz, A. H., & Kjell, O. N. E. (in progress). *Mental health recommendations from natural language responses closely align with best-estimate expert assessments: A sequential evaluation with model pre-registration*.

Winter, D. G. (1991). Measuring personality at a distance: Development of an integrated system for scoring motives in running text. In D. J. Ozer, J. M. Healy, Jr., & A. J. Stewart (Eds.), *Perspectives in personality, Vol. 3. Part A: Self and emotion; Part B: Approaches to understanding lives* (pp. 59–89). Jessica Kingsley Publishers.

Yeung, R. C., Danckert, J., Van Tilburg, W. A., & Fernandes, M. A. (2024). Disentangling boredom from depression using the phenomenology and content of involuntary autobiographical memories. *Scientific Reports*, *14*(1), 2106.

Zhou, S., Wang, N., Wang, L., Liu, H., & Zhang, R. (2022). CancerBERT: A cancer domain-specific language model for extracting breast cancer phenotypes from electronic health records. *Journal of the American Medical Informatics Association*, *29*(7), 1208–1216.

Zhou, W., Prater, L. C., Goldstein, E. V., & Mooney, S. J. (2023). Identifying rare circumstances preceding female firearm suicides: Validating a large language model approach. *JMIR Mental Health*, *10*(1), Article e49359. https://doi.org/10.2196/49359

Zimmermann, J., Brockmeyer, T., Hunn, M., Schauenburg, H., & Wolf, M. (2017). First-person pronoun use in spoken language as a predictor of future depressive symptoms: Preliminary evidence from a clinical sample of depressed patients. *Clinical Psychology & Psychotherapy*, *24*(2), 384–391.